



THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE ■

Switching regressions with imperfect regime classification information: theory and applications

LSE Research Online URL for this paper: <http://eprints.lse.ac.uk/103119/>

Version: Published Version

Monograph:

Hajivassiliou, Vassilis (2019) Switching regressions with imperfect regime classification information: theory and applications. STICERD Econometrics Papers (610). Suntory and Toyota International Centres for Economics and Related Disciplines, London, UK.

Reuse

Items deposited in LSE Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the LSE Research Online record for the item.

Switching Regressions with Imperfect Regime Classification Information: Theory and Applications

by

Vassilis A. Hajivassiliou*

Department of Economics, STICERD, and Financial Markets Group
London School of Economics

November 2019

Extended Abstract

This paper discusses switching regressions econometric modelling with imperfect regime classification information. The econometric novelty is that misclassification probabilities are allowed to vary endogenously over time. Standard maximum likelihood estimation is infeasible in this case because each likelihood contribution requires the evaluation of 2^T terms (where T is the number of observations available). We develop an algorithm that allows efficient estimation when such imperfect information is available, by evaluating the exact likelihood through simply T matrix multiplications (each of a 2×2 matrix times a 2×1 vector.) Our methods are shown to be widely applicable to various areas of economic analysis such as to Hamilton's work on Markov-Switching models in Macroeconomics; to external financing problems faced by firms in Corporate Finance; and to game-theoretic models of price collusion in Industrial Organization.

We proceed to apply our methods to analyze price fixing by the Joint Executive Committee railroad cartel from 1880 to 1886 and develop tests of two prototypical game-theoretic models of tacit collusion. The first model, due to Abreu, Pearce and Stacchetti (1986), predicts that price will switch across regimes according to a Markov process. The second model, by Rotemberg and Saloner (1986), concludes that price wars are more likely in periods of high industry demand. Switching regressions are used to model the firms' shifting between collusive and punishment behaviour. The JEC data set is expanded to include measures of grain production to be shipped and availability of substitute transportation services. Our findings cast doubt on the applicability of the Rotemberg and Saloner model to the JEC railroad cartel, while they confirm the Markovian prediction of the Abreu et al. model.

Keywords: Switching regressions models, Measurement Errors, Trigger-price mechanisms, Price-fixing

JEL Classifications: C72, L12, C51, C52, C15

*This research has been partly financed by NSF #SES 8813630. I would like to thank David Roth for expert research assistance and for helpful discussions, and Rosemarie Lewis for editing the manuscript. I have benefitted from constructive comments by Jim Heckman, Yannis Ioannides, Al Klevorick, Lung-Fei Lee, Rick Levin, Nancy Lutz, Ariel Pakes, David Pearce, Julio Rotemberg, Anne Royalty, Mike Whinston, and by participants at seminars at Harvard/M.I.T., Yale, Berkeley, Stanford, Virginia, Munich, and University College, London. I thank in particular Jean-Pierre Florens for highlighting the relevance of my algorithm to Hamilton's work on Markov-switching macroeconomic models. The usual disclaimer applies.

1 Introduction

This paper discusses switching regressions econometric modelling with imperfect regime classification information. The econometric novelty is that misclassification probabilities are allowed to vary endogenously over time. Standard maximum likelihood estimation is infeasible in this case because each likelihood contribution requires the evaluation of 2^T terms (where T is the number of observations available). We develop an algorithm that allows efficient estimation when such imperfect information is available, by evaluating the exact likelihood through simply T matrix multiplications (each of a 2×2 matrix times a 2×1 vector.) Our methods are shown to be widely applicable to various areas of analysis such as to Hamilton's work on Markov-Switching models in Macroeconomics (1996a; 1996b; 1996c; 2001; 2002); to external financing problems faced by firms in Corporate Finance (Hajivassiliou and Savignac (2019)); and to game-theoretic models of price collusion in Industrial Organization. (See Section 7 below.)

Section 2 analyzes problems of estimation when there are *measurement errors in regime-classification information*. In this case, the ML estimator that treats imperfect classifying information as perfect is inconsistent (Lee and Porter (1984)). Moreover, ML estimators that do not use regime-classifying information are in general either seriously inefficient (Goldfeld and Quandt (1975)) or not identified. The misclassification problem is ubiquitous in all econometric applications of the switching-regression methodology.¹ I also examine estimators that incorporate appropriately imperfect classification information in the form of (multiple) indicator variables. A major difference between my procedure and the Lee and Porter (1984) analysis is that mine allows the probabilities of misclassification *to vary* over the sample period and also *be endogenously determined*. These features are expected to be crucial once exogenous classifying information is available, because it is normally harder to accurately classify a market when it is closer to a transition period.²

Section 3 discusses estimation methods for switching models with imperfect classification information when switching occurs according to a Markov process. The problem of classical *errors in measuring explanatory variables* is also encountered in our empirical application and is studied in Section 4. The identification of the econometric model with varying misclassification probabilities is established in Section 5.

¹Early such applications include Lee and Porter (1984) for an analysis of cartel stability and Lee (1978) for an analysis of unionization.

²The methodology is in the spirit of the Tobit model of Nelson (1977) with a stochastically unobservable threshold.

The indices for industry demand and for extra-cartel competition constructed for this paper may contain serious measurement errors; hence, it is important that estimation methods allow for this possibility.³ In Section 6 we derive a recursion relation that makes tractable the evaluation of the likelihood functions of our general econometric models.

We proceed in Section 7 to apply our methods to analyze price fixing by the Joint Executive Committee railroad cartel from 1880 to 1886 and develop tests of two prototypical game-theoretic models of tacit collusion. The first model, due to Abreu, Pearce and Stacchetti (1986), predicts that price will switch across regimes according to a Markov process. The second model, by Rotemberg and Saloner (1986), concludes that price wars are more likely in periods of high industry demand. Switching regressions are used to model the firms' shifting between collusive and punishment behaviour. The JEC data set is expanded to include measures of grain production to be shipped and availability of substitute transportation services. The findings cast doubt on the applicability of the Rotemberg and Saloner model to the JEC railroad cartel, while they confirm the Markovian prediction of the Abreu et al. model.

Section 8 concludes.

2 Imperfect Classification Information in Switching Regressions

The econometric methods we present in this Section have wide applicability in various areas of economic analysis. Examples are: Hamilton's work on Markov-Switching models in Macroeconomics (1996a; 1996b; 1996c; 2001; 2002); external financing problems faced by firms in Corporate Finance (Hajivassiliou and Savignac (2019)); and game-theoretic models of price collusion in Industrial Organization (see Section ?? below).

2.1 A Single Regime Indicator

Consider the general switching-regression model:

$$y_{it}^* = h_i(X_i \delta_i) + \epsilon_{it} \quad i = 0, 1; \quad t = 1, \dots, T \quad (2.1.a)$$

$$y_{2t}^* = h_2(Z_t \zeta) + \epsilon_{2t} \quad (2.1.b)$$

³We show in that Section how simulation estimation methods (McFadden (1989), Pakes and Pollard (1989)) can handle the concomitant high-dimensional integrals that arise in nonlinear errors-in-variables models such as the switching-regression models of this paper.

$$y_{3t}^* = y_{2t}^* + \eta_t. \quad (2.1.c)$$

Here, y_{0t}^* , y_{1t}^* , y_{2t}^* , and y_{3t}^* are latent variables, unobservable by the econometrician; X_0 , X_1 , X_2 , and Z are matrices of explanatory (exogenous) variables; and $(\epsilon_{0t}, \epsilon_{1t}, \epsilon_{2t}, \eta_t)'$ is multivariate normally distributed, *i.i.d.* over time, with zero-mean. The functions $h_i(\cdot)$, $i = 0, 1, 2$, are known to the econometrician up to the vectors of parameters δ_i and ζ , which will be estimated.

The econometrician observes the (endogenous) variable Y_t , which is generated as follows:

$$Y_t = \begin{cases} y_{1t}^* & \text{iff } y_{2t}^* \geq 0 \\ y_{0t}^* & \text{otherwise} \end{cases} \quad (2.2)$$

In standard terminology, the two equations (2.1.a), $i = 0, 1$, are termed the “switched” equations and (2.1.b) the “switching” equation. Using the indicator function introduced above, we define the dummy variables $I_t \equiv \mathbf{1}(y_{2t}^* \geq 0)$ and $D_t \equiv \mathbf{1}(y_{3t}^* \geq 0)$. The econometrician observes D_t but not I_t . As long as $\sigma_\eta^2 > 0$, D_t is an imperfect measurement of I_t . In this sense, η_t can be thought of as *coding error*.

In its general form without measurement errors in regime classification, the switching-regression model was used by Lee (1978) to study union/nonunion wage determination.⁴ As Lee and Porter (1984) explain, using inaccurate regime classification information in ML estimation leads to inconsistency. Moreover, Goldfeld and Quandt (1975) show that if perfect information is not used, ML estimation is seriously inefficient. Hajivassiliou (1987) combines these results to derive Hausman (1978) tests of accuracy of classification information.

Lee and Porter (1984) allowed for a *constant* probability that observations were misclassified into the two regimes; their only explanatory variable in the switching equation, Z , was a constant. But assuming a constant probability of misclassification is inappropriate if one expects the probability of misclassification to vary over time, and especially so if one has exogenous information represented by Z_t , which, as theory suggests, should affect switching.

I model the misclassification probability as a monotonic function of the (unobservable) propensity of the industry to be in a particular regime measured by the latent variable y_{2t}^* . For example, in the disequilibrium version of the switching model (Fair and Jaffee (1972)), it seems plausible to assume that the probability of misclassification is smaller the larger the level of excess demand in the system. I demonstrate shortly that the coding error equation (2.1.c) incorporates this property into the

⁴Fair and Jaffee (1972), *inter alia*, used the model to analyze markets in disequilibrium, by letting y_{1t}^* denote notional demand in period t , y_{0t}^* notional supply, and y_{2t}^* excess demand.

model.

The contribution of an (independent) observation t to the likelihood function of the switching-regression model with coding error can be derived as follows: First observe that

$$\begin{aligned} \text{for } D_t = 1 : \quad & y_{3t}^* \geq 0 \quad \text{if } y_{2t}^* \geq 0, \eta_t \geq -y_{2t}^* \quad Y_t = y_{1t}^* \quad (I_t = 1) \\ & \text{if } y_{2t}^* < 0, \eta_t \geq -y_{2t}^* \quad Y_t = y_{0t}^* \quad (I_t = 0) \end{aligned} \quad (2.3)$$

$$\begin{aligned} \text{for } D_t = 0 : \quad & y_{3t}^* < 0 \quad \text{if } y_{2t}^* < 0, \eta_t < -y_{2t}^* \quad Y_t = y_{0t}^* \quad (I_t = 0) \\ & \text{if } y_{2t}^* \geq 0, \eta_t < -y_{2t}^* \quad Y_t = y_{1t}^* \quad (I_t = 1) \end{aligned}$$

Let us use the notation $p_{d|i} \equiv \text{prob}(D_t = d | I_t = i)$, $p_{di} \equiv \text{prob}(D_t = d, I_t = i)$, $p_{dt} = \text{prob}(D_t = d)$, $\pi_{it} = \text{prob}(I_t = i)$, and $f_{it} = \text{pdf}(y_{it}^*)$, where d and i take values 0 or 1. For simplicity assume that ϵ_{0t} and ϵ_{1t} are independent of ϵ_{2t} and η_t .⁵ Dropping the t subscript for simplicity, this specification implies that the log-likelihood contribution is:

$$\text{prob}(D, y | X) = D \cdot \ln(p_{1|1}f_1 + p_{1|0}f_0) + (1 - D) \cdot \ln(p_{0|1}f_1 + p_{0|0}f_0). \quad (2.4)$$

Note that the p_{di} 's involve bivariate integrals of the form

$$p_{d|i} = \int \int_{S_{DI}} f(\epsilon_2, \theta) d\epsilon_2 d\theta / \int_{S_I} f(\epsilon_2) d\epsilon_2, \quad (2.5)$$

where $\theta \equiv \epsilon_2 - \eta$, and the regions of integration (as described in (2.3)) are the sets:

$$S_{DI} = \{\epsilon_2 \stackrel{>}{I} - Z\zeta, \eta \stackrel{>}{D} - (Z\zeta + \epsilon_2)\} \text{ and } S_I = \{\epsilon_2 \stackrel{>}{I} - Z\zeta\},$$

$$\text{where } \stackrel{>}{I} \equiv \{\geq \text{ if } I = 1, < \text{ if } I = 0\} \text{ and } \stackrel{>}{D} \equiv \{\geq \text{ if } D = 1, < \text{ if } D = 0\}.$$

The common distributional assumption of normality is imposed.

The coding error model with the likelihood function defined by using (2.3)–(2.5) possesses the desired property that the misclassification probability is highest at the borderline case when a regime switch appears most likely, and falls monotonically as the exogenous classifying information becomes stronger. To see this, first note that the probabilities of misclassification are:

$$\begin{aligned} (D = 1 | I = 0) : \quad & p_{1|0} = \text{prob}(\eta_t \geq -y_{2t}^* | y_{2t}^* < 0) \\ (D = 0 | I = 1) : \quad & p_{0|1} = \text{prob}(\eta_t < -y_{2t}^* | y_{2t}^* \geq 0) \end{aligned} \quad (2.6)$$

Figure 1 presents probability plots for the misclassification case of $D = 1$ and $I =$

⁵This assumption can be relaxed at the cost of further computational complexity.

0 as a function of the exogenous part of the switching equation, $Z\zeta$.⁶ Various values of the standard deviation of the coding error η are considered. As can be seen from Figure 1a, the conditional probability of misclassification, $p_{1|0}$, is monotonic in $Z\zeta$ in the desired direction, rising when the signal $Z\zeta$ tends to suggest the wrong regime more strongly. For example, when the true state of the system is no collusion ($I = 0$), higher values of $Z\zeta$ are further at odds with the truth, hence $\text{Prob}(D = 1|I = 0)$ rises. As the standard deviation of the coding error η rises, the signal becomes less informative; in the limit, when $\sigma_\eta \rightarrow \infty$, the misclassification probabilities ($\text{Prob}(D = d|I = i), d \neq i$) approach 0.5. Hence, we confirm that the switching model with coding error introduced here possesses the desired property that the misclassification probability falls as the tendency to lie in a particular regime rises. In Figure 1b we see that the joint probability of misclassification p_{10} has a unique mode at the least informative value of the signal, $Z\zeta = 0$, since in such a case it is most difficult to correctly classify the particular period.

***** FIGURES 1a, 1b, 1c about HERE *****

An important caveat is that the coding-error switching-regression model allows only a limited degree of systematic misclassification. For example, despite the presence of the coding errors, the only change in the discrete part of the model, (2.1.b), is in the variance of the latent variable y_{2t}^* , which is, of course, unidentified. This is illustrated in Figure 1c. Hence, one can obtain consistent estimates for ζ up to scale despite such misclassification.⁷ This, however, does not imply that the presence of the coding error is unimportant, because ML estimation of the complete discrete/continuous switching-regression model would still yield inconsistent results if the measurement errors were neglected.

2.2 Multiple Regime Indicators

Finally, suppose we have M multiple indicators D_1, \dots, D_M of regime classification. This is the nonlinear analogue of the classic MIMIC model of Goldberger (1972). We then obtain 2^{M+1} categories with respect to D_1, \dots, D_M , and I .⁸

The likelihood contributions will in general involve $(M + 1)$ -fold integrals, which can be calculated by numerical methods for M up to 2 or 3. This modelling approach, like the coding-error model with a single indicator, (2.1)–(2.2), also has the

⁶The corresponding plots for the case with $D = 0$ and $I = 1$ are exact mirror images with respect to $Z\zeta = 0$ of those in Figure 1 and are not given separately.

⁷The importance of this restrictive feature of my measurement errors model will be investigated in future work.

⁸For the purposes of the empirical implementation in Section 7.4 with two imperfect classification indicators, I define $R \equiv Z\zeta$ and give the eight possibilities in that case:

desirable property that the misclassification probabilities vary over the sample period depending on the true probability of switching.⁹

3 A Markovian Switching Model with Imperfect Classification

Because of the *i.i.d.* assumptions on the error vector $(\epsilon_{0t}, \epsilon_{1t}, \epsilon_{2t}, \eta_t)'$, the models of the previous section exhibit a Bernoulli switching structure, conditional on the exogenous variables. This is characterized by a transition matrix:

$$\begin{array}{ccc} & I_t = 1 & I_t = 0 \\ \begin{array}{l} I_{t-1} = 1 \\ I_{t-1} = 0 \end{array} & \begin{array}{cc} \tau_t & 1 - \tau_t \end{array} & \begin{array}{c} \\ \end{array} \quad \text{Bernoulli} \end{array} \quad (3.1)$$

In (3.1) the transition probabilities τ 's depend on time only through the exogenous variables, but not on the past state variable. Next I introduce a model that allows the switching process to exhibit Markov dependence over time. This is necessary to test the key prediction of Markovian switching of the game-theoretic model of Abreu et al. (1986).

If I_t is a Markov process, then it has the transition structure:

$$\begin{array}{ccc} & I_t = 1 & I_t = 0 \\ \begin{array}{l} I_{t-1} = 1 \\ I_{t-1} = 0 \end{array} & \begin{array}{cc} \tau_{11t} & 1 - \tau_{11t} \\ \tau_{01t} & 1 - \tau_{01t} \end{array} & \begin{array}{c} \\ \end{array} \quad \text{Markov} \end{array} \quad (3.2)$$

where $\tau_{ijt} = \text{Prob}(I_t = i | I_{t-1} = j)$.¹⁰ Specifically, to introduce a Markov structure

D_1	D_2	I	$\epsilon_2 - \eta_1$	$\epsilon_2 - \eta_2$	ϵ_2
1	1	1	$\epsilon_2 - \eta_1 > 0$	$\epsilon_2 - \eta_2 > 0$	$\epsilon_2 \leq R$
1	1	0	$\epsilon_2 - \eta_1 > 0$	$\epsilon_2 - \eta_2 > 0$	$\epsilon_2 > R$
1	0	1	$\epsilon_2 - \eta_1 > 0$	$\epsilon_2 - \eta_2 \leq 0$	$\epsilon_2 \leq R$
1	0	0	$\epsilon_2 - \eta_1 > 0$	$\epsilon_2 - \eta_2 \leq 0$	$\epsilon_2 > R$
0	1	1	$\epsilon_2 - \eta_1 \leq$	$\epsilon_2 - \eta_2 > 0$	$\epsilon_2 \leq R$
0	1	0	$\epsilon_2 - \eta_1 \leq$	$\epsilon_2 - \eta_2 > 0$	$\epsilon_2 > R$
0	0	1	$\epsilon_2 - \eta_1 \leq$	$\epsilon_2 - \eta_2 \leq 0$	$\epsilon_2 \leq R$
0	0	0	$\epsilon_2 - \eta_1 \leq$	$\epsilon_2 - \eta_2 \leq 0$	$\epsilon_2 > R$

⁹Under the normality assumptions discussed above, in the empirical implementations below the bivariate integrals are calculated through an algorithm of Divgi (1979). In the case of two indicator variables in Section 7.4, the implied trivariate integrals are calculated through the method in Steck (1958). The higher dimension integrals implied by more indicators can be accommodated by simulation estimation methods. See Hajivassiliou (1993) for a discussion.

¹⁰One expects positive serial persistence, in the sense of $\tau_{11t} > \tau_{10t}$.

of order 1, I modify the switching equation (2.1.b) so that the propensity to switch, y_{2t}^* , depends on the lagged state I_{t-1} , i.e.,

$$I_t = \begin{cases} 1 & \text{if } Z_t\zeta + \rho I_{t-1} + \epsilon_{2t} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3.3)$$

With perfect classification information, this structure is straightforward to estimate since¹¹

$$\begin{aligned} p[\mathbf{Y}, \mathbf{I}, I_0 | \mathbf{X}] &\equiv p[Y_1, \dots, Y_T, I_1, \dots, I_T, I_0 | X_1, \dots, X_T] \\ &= p[Y_T, I_T | I_{T-1}, X_T] \cdot p[Y_{T-1}, I_{T-1} | I_{T-2}, X_{T-1}] \cdots p[Y_2, I_2 | I_1, X_2] \cdot p[Y_1, I_1 | I_0, X_1] \cdot p[I_0]. \end{aligned} \quad (3.4)$$

The likelihood function for process (3.2)–(3.3), however, becomes extremely intractable in the presence of imperfect regime-classification information because it then requires the evaluation of 2^T terms. The reason is as follows. We can readily show that

$$p[D_t, I_t | I_{t-1}] =$$

$$I_t p[D_t, I_t = 1 | I_{t-1}] + (1 - I_t) p[D_t, I_t = 0 | I_{t-1}] \quad (3.5.a)$$

$$p[Y_t, D_t | I_{t-1}] =$$

$$D_t [f_1 I_t p[1, 1 | I_{t-1}] + f_0 (1 - I_t) p[1, 0 | I_{t-1}]] + (1 - D_t) [f_1 I_t p[0, 1 | I_{t-1}] + f_0 (1 - I_t) p[0, 0 | I_{t-1}]], \quad (3.5.b)$$

where I_t is determined by (3.3). But the econometrician only observes D_t , given by

$$D_t = \begin{cases} 1 & \text{if } Z_t\zeta + \rho I_{t-1} + \epsilon_{2t} + \eta_t \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3.6)$$

Since I_{t-1} is unobserved by the econometrician for all t , the likelihood function is

$$p[\mathbf{Y}, \mathbf{D} | \mathbf{X}] = \sum_{I_T} \sum_{I_{T-1}} \cdots \sum_{I_2} \sum_{I_1} \sum_{I_0} p[Y_T, D_t, I_T | I_{T-1}] \cdots p[Y_1, d_1, I_1 | I_0] \cdot p[I_0]. \quad (3.7)$$

Because each pair of consecutive terms involves I_{t-1} , the likelihood $p[\mathbf{Y}, \mathbf{D} | \mathbf{X}]$ will in general require the evaluation of 2^T terms, a patently intractable task when T is of the order of 300, as in this paper. To solve this problem I show in Section 6 that, by extending ideas in Cosslett and Lee (1985) and Moran (1986), a recursion relation can be derived that makes evaluation of (3.7) feasible.

¹¹Note that it is not crucial how one treats $p[I_0]$, since this term has asymptotically vanishing influence. This is in contrast to the longitudinal data set case.

Note again that the approach here differs fundamentally from that of Lee and Porter (1984) and Cosslett and Lee (1985) in that the probability of misclassification is not constant but varies monotonically with the magnitude of $Z_t\zeta + \rho I_{t-1}$. A priori, this is a realistic feature. Given the dependence over time described in (3.2), one should expect the probability of misclassification to vary over time; it should be highest close to the boundary points when a switch occurs. These properties are exhibited by the conditional probability expressions above.¹²

4 Classical Measurement Errors

At least two explanatory variables that I shall use to test the game-theoretic models are measured with potentially serious errors. The variables most suspect are the constructed index of grain production to be shipped and the constructed measure of the availability and strength of extra-cartel competition. Hence, in this section I investigate the effect of measurement errors in the explanatory variables of nonlinear models of the type estimated here. I will show that structural ML estimation of such models introduces high-dimensional integration problems, which can be avoided by employing simulation estimation methods (McFadden (1989), Pakes and Pollard (1989)).

To see how multiple integrals arise in models we analyze in this paper, consider the general limited dependent variable (LDV) model

$$y_t^* = z_t^* \beta + \epsilon_t, \quad (4.1)$$

where the econometrician observes $y_t = \tau(y_t^*)$. The function $\tau(\cdot)$ maps the vector of underlying latent variables, y_t^* , into the vector of observable endogenous variables, y_t . For example, in the switching-regression model of Section 2, $y_t^* = (y_{0t}^*, y_{1t}^*, y_{2t}^*, y_{3t}^*)'$, $y_t = (Y_t, D_t)$, and the $\tau(\cdot)$ function is specified implicitly in equation (2.3). In (4.1), z_t^* is a $k \times 1$ vector of explanatory variables that are not directly observed. Instead we have the imperfect measurement x_t , given by

$$x_t = z_t^* \delta + v_t. \quad (4.2)$$

In terms of the observable limited dependent variable y_t , the model (4.1) can be

¹²There is a cost, however, in terms of computational complexity because the conditional probability expressions $p[D_t, I_t | I_{t-1}]$ now involve bivariate normal integrals (and in general $(M+1)$ -fold integrals when M imperfect regime indicator variables are available).

written as:

$$y_t = \tau(z_t^* \beta + \epsilon_t). \quad (4.3)$$

Assuming that the measurement errors are of the classical form, it is plausible to postulate that $f(y|z^*, x) = f(y|z^*)$, which follows from mean-independence of the measurement error v_t and the true variables z_t^* .¹³ Hence, using basic properties of conditional probability functions, the likelihood contribution conditional on the observable vector x is

$$\begin{aligned} f(y|x) &= \int f(y, z^*|x) dz^* = \int f(y, z^*, x) / f(x) dz^* = \\ &= \int f(y|z^*, x) \cdot f(z^*|x) dz^* = \int f(y|z^*) \cdot f(z^*|x) dz^*. \end{aligned} \quad (4.4)$$

Equation (4.4) illustrates that integration of order equal to the number of x variables will be needed to evaluate the likelihood in terms of the observable variables y and x . This is *in addition* to any integration required to calculate the density of y conditional on the unobservable z^* variables. These difficulties arise because the mismeasured regressors appear inside the nonlinear function $\tau(\cdot)$. For example, in the switching-regression models of the previous sections, one must first calculate the likelihood assuming all explanatory variables are observed without error, $f(y|z^*)$. Then double numerical quadrature is required to evaluate $f(y|x)$, given that two of the z^* 's are observed imperfectly.¹⁴ In the estimation section below, I offer both quadrature-based ML estimates with two mismeasured explanatory variables, as well as estimates by simulation estimation methods. This allows a comparative evaluation of the latter.¹⁵

¹³An alternative way of introducing measurement errors in the explanatory variables is through *mixture models*. This approach simultaneously estimates non-parametrically the (x, z^*) relation. Given the relatively small sample size for my complicated nonlinear model, I chose instead to parameterize explicitly the measurement errors to be of the classical type. For a general review of measurement error models, see Fuller (1987).

¹⁴A second difficulty that arises in nonlinear models with classical measurement errors in RHS variables is that the covariance matrix of the $f(y|z, x)$ distribution is a general $k \times k$ matrix Σ , even when ϵ_t has a scalar covariance matrix. This problem is explained in the Online Extended version of this paper.

¹⁵The method employed here is Maximum Simulated Likelihood (MSL) based on the Smooth Recursive Conditioning simulator (SRC) of Börsch-Supan and Hajivassiliou (1993). For a more detailed discussion of this method as well as of other simulation estimation methods for limited dependent variable models that are continuous in the parameter vector, see Hajivassiliou (1994) and Hajivassiliou and McFadden (1998).

5 Identification of Switching Model with Imperfect Regime Classification

In this Section I show how all the parameters of the switching-regression model with coding error (2.1.a)–(2.1.c) are econometrically identified, subject to the normalization that $\sigma_2 = 1$. Recall the definitions

$$p_{di} \equiv \text{Prob}(D = d, I = i) \quad (5.1.a)$$

$$p_{d|i} \equiv \text{Prob}(D = d|I = i) \quad (5.1.b)$$

$$p_d \equiv \text{Prob}(D = d) \quad (5.1.c)$$

$$\pi_i \equiv \text{Prob}(I = i). \quad (5.1.d)$$

Under the normality assumptions imposed and the normalization $\sigma_2 = 1$

$$\begin{aligned} p_d &= D \cdot \text{Prob}(D = 1) + (1 - D) \cdot \text{Prob}(D = 0) = \\ &D \cdot \Phi(Z\zeta/\sqrt{1 + \sigma_\eta^2}) + (1 - D) \cdot (1 - \Phi(Z\zeta/\sqrt{1 + \sigma_\eta^2})) \end{aligned} \quad (5.2)$$

$$\pi_i = I \cdot \text{Prob}(I = 1) + (1 - I) \cdot \text{Prob}(I = 0) = I \cdot \Phi(Z\zeta) + (1 - I) \cdot (1 - \Phi(Z\zeta)) \quad (5.3)$$

Using only the imperfect classification indicator D , from the marginal likelihood (5.2) we can estimate the expression $\zeta/\sqrt{1 + \sigma_\eta^2}$ consistently.

Now consider the marginal likelihood for the observed endogenous variable y , neglecting any classification information, i.e., consider the marginal likelihood

$$f(y) = f_1 \cdot \pi_1 + f_0 \cdot \pi_0 = f_1 \cdot \Phi(Z\zeta) + f_0 \cdot (1 - \Phi(Z\zeta)). \quad (5.4)$$

From this, we can obtain consistent estimates for the parameters δ_1 , δ_0 , σ_1 , σ_0 , and ζ , provided either the functions $h_0(\cdot)$ and $h_1(\cdot)$ are not the same, or further restrictions on δ_1 and δ_0 are imposed.

Finally, consider the conditional likelihood

$$f(y|D = 1) = f_1 \cdot \frac{p_{11}}{p_1} + f_0 \cdot \frac{p_{10}}{p_1}, \quad (5.5)$$

which uses separately the observations classified by the (imperfect) indicator D to be in collusion. We see immediately that in such a case the expressions p_{11}/p_1 and

p_{10}/p_1 are consistently estimable. But

$$p_{11}/p_1 = p_{1|1} \cdot \pi_1/p_1 = p_{1|1} \cdot \Phi(Z\zeta)/\Phi(Z\zeta/\sqrt{1 + \sigma_\eta^2}); \quad (5.6)$$

hence, $p_{1|1}$ can also be identified. The identification of the remaining term $p_{0|0}$ follows from exactly analogous arguments.

6 A Recursion Algorithm for the Markovian Switching-Regression Model with Coding Error

The aim is to facilitate evaluation of the likelihood function of Section 3, which is given by:

$$p[\mathbf{y}, \mathbf{D}|X] = \sum_{I_T} \sum_{I_{T-1}} \cdots \sum_{I_2} \sum_{I_1} \sum_{I_0} p[y_T, D_t, I_T|I_{T-1}] \cdots p[y_1, d_1, I_1|I_0] \cdot p[I_0]. \quad (6.1)$$

The difficulty in evaluating (6.1) directly is that each pair of consecutive terms involves I_{t-1} ; hence, each likelihood evaluation will require calculating 2^T terms, which is a computationally prohibitive task.

The following arguments generalize ideas in Cosslett and Lee (1985) and Moran (1986) and show how (6.1) can be evaluated recursively through T matrix multiplications. Define the set of available endogenous information at time t by S_t , i.e., $S_t \equiv (y_1, D_1, y_2, D_2, \dots, y_t, D_t)$. Further define $Q_t(I_t) \equiv p[S_t, I_t]$. Since we can always write

$$Q_t(I_t) = p[S_{t-1}, y_t, D_t, I_t] = \sum_{I_{t-1}} p[S_{t-1}, I_{t-1}, y_t, D_t, I_t], \quad (6.2)$$

it follows that

$$Q_t(I_t) = \sum_{I_{t-1}} p[y_t, D_t, I_t|I_{t-1}, S_{t-1}] \cdot p[I_{t-1}, S_{t-1}] = \sum_{I_{t-1}} p[y_t, D_t, I_t|I_{t-1}] \cdot Q_{t-1}(I_{t-1}), \quad (6.3)$$

where we have used the Markov structure $p[y_t, D_t, I_t|I_{t-1}, S_{t-1}] = p[y_t, D_t, I_t|I_{t-1}]$ and the definition $Q_{t-1}(I_{t-1}) \equiv p[I_{t-1}, S_{t-1}]$. But calculation of (6.3) only requires information up to t , as the following matrix equation shows:

$$\begin{pmatrix} Q_T(0) \\ Q_T(1) \end{pmatrix} =$$

$$\begin{pmatrix} p[y_t, D_t, I_t = 0 | I_{t-1} = 0] & p[y_t, D_t, I_t = 0 | I_{t-1} = 1] \\ p[y_t, D_t, I_t = 1 | I_{t-1} = 0] & p[y_t, D_t, I_t = 1 | I_{t-1} = 1] \end{pmatrix} \cdot \begin{pmatrix} Q_{T-1}(0) \\ Q_{T-1}(1) \end{pmatrix} \quad (6.4)$$

or,

$$\mathbf{Q}_t = \mathbf{M}_t \cdot \mathbf{Q}_{t-1}.$$

The likelihood (6.1) can thus be calculated recursively from (6.4) and

$$p[\mathbf{y}, \mathbf{D} | X] = \sum_{I_T} Q_T(I_T) = Q_T(0) + Q_T(1). \quad (6.1')$$

7 Empirical Application: Testing Prototypical Models of Price Fixing

We now use the methods developed above to test two prototypical game-theoretic models of cartel behaviour. The first model, due to Abreu, Pearce, and Stacchetti (1986), predicts that behaviour switches across collusive/price-warfare regimes according to a Markov process. The second model, due to Rotemberg and Saloner (1986), predicts that the probability of a price war is higher in periods of high industry demand. Maximum likelihood (ML) and simulation estimation methods that allow for measurement errors in switching-regression models are presented and applied to test these models using data on the Joint Executive Committee (JEC) railroad cartel for the period 1880-1886.

Subsection 7.1 discusses the Abreu, Pearce, and Stacchetti (1986) model [APS] and the Rotemberg and Saloner (1986) model [RS]. It should be noted that the first game-theoretic model of tacit collusion to predict switching behaviour across collusive and price-warfare regimes was the model of Green and Porter (1984), which has already been tested elsewhere. (See Lee and Porter (1984) and Porter (1983b)). We then present a simple model of a symmetric oligopoly which nests the basic predictions of APS and RS. This second prediction seems counter to the conventional view of the classical industrial organization literature (see for example Stigler (1964)). The econometric framework developed in subsection 7.2 allows one to employ imperfect regime-classification information, which is available from several sources. The use of regime-classification information is a novel feature of the paper that distinguishes it from Berry and Briggs (1988), who tested the Markovian prediction of the Abreu et al. model, but did not exploit such information.

The JEC data used in this study are discussed in subsection 7.3. The data set, originally developed by Porter (1983b) and Lee and Porter (1984), is expanded to include measures of grain production to be shipped and the availability of substitute

transportation services. The construction of these measures is described in Data Appendix.

Section 7.4 presents the empirical implementation. We find that the predictions of the RS model are not borne out by this data set. The evidence favours instead the Markovian prediction of the APS theory.

7.1 Prototypical Game-Theoretic Models of Price Fixing Behaviour

Lee and Porter (1984) and Porter (1983b) used switching-regression methodology to test the game-theoretic models of Porter (1983a) and Green and Porter (1984). In this paper I consider two other models of price fixing, one by Rotemberg and Saloner (1986) and the second by Abreu et al. (1986).

I first summarize the Rotemberg and Saloner (1986) model.¹⁶ Consider a symmetric n -firm, price-setting cartel facing stochastic demand. At each period the level of demand is a random variable independently and identically distributed (*i.i.d.*) over time. The firms learn the realized state of demand *before* making (simultaneously) their price choices. When demand is high, each firm feels a temptation to undercut its competitors in order to take advantage of high demand now, because it does not expect it to persist. The *i.i.d.* assumption leads firms to expect that demand will be lower next period. Hence, a punishment by the competitors would appear less severe than if firms believed it likely that high demand would persist in the next period. As a result, Rotemberg and Saloner (1986) predict that in the presence of observable demand shocks, price wars (in the sense of less collusion) will occur mostly during industry booms.¹⁷ This prediction is contrary to the conventional wisdom of the traditional industrial organization literature, which holds that it is generally more difficult to collude successfully during industry recessions when each firm is possibly preoccupied with its own survival. Rotemberg and Saloner (1986), pp.395-396, describe how modification of their basic model yields behaviour that fluctuates between periods of cooperation and non-cooperation. This arises when the strategy space is restricted so that the oligopoly can choose only between the joint monopoly and the competitive prices. They believe this version of their model is “intuitively more appealing,” and they show that the basic prediction that price wars are more likely to occur in demand booms is preserved.

The alternative game-theoretic model of collusion I shall test is due to Abreu et al.

¹⁶For an excellent review of game-theoretic models of tacit collusion see Chapter 6 in Tirole (1988).

¹⁷These are not necessarily price wars in the usual sense of periods of maximal punishment of Bertrand (competitive) behaviour, because the price may actually be higher during booms than during recessions.

(1986). In this model, firms do not observe their competitors' quantities but rather the market price whose distribution is determined by industry output and realized demand. Demand shocks are *i.i.d.* over time but are not observed by the firms. In this supergame, the firms have concave objective functions, and the distribution of the market price p_t conditional on aggregate output Q_t is assumed to have the property that a low price is more likely to have arisen from a high Q_t than from a low one. Abreu et al. (1986) are able to show that under these assumptions price wars will result; the behaviour of firms will be characterized by a trigger scheme, usually a "tail test." During periods of successful collusion, each firm will be producing q^+ and earning payoff V^+ . A trigger-price level p^+ will be determined such that observation of a price lower than p^+ will trigger a punishment phase in which firms will each produce a (higher) output q^- and earn a lower payoff V^- . This provides incentives for firms to restrict output. A second trigger p^- will determine whether a punishment phase will persist or whether the industry will revert to successful collusion. Since a harsh punishment requires a high output, reversion to successful collusion will involve an "inverse tail test": if a high price (greater than p^-) is observed, the game remains in the punishment phase. Conversely, successful collusion resumes when a price lower than the p^- threshold is observed; again, this provides incentives for high output.

***** FIGURE 2 about HERE *****

Figure 2 shows two examples of the distribution of price given total output $f(p_t|Q_t)$, one when aggregate output is collusive (Q^+) and one when the industry is producing high output (Q^-); the trigger prices p^+ and p^- also appear in the figure. Define the successful collusion indicator I_t such that $I_t = 1$ indicates collusive output is produced in period t and $I_t = 0$ indicates that this is a punishment period. The Abreu et al. model predicts that industry behaviour switches between periods of successful collusion and punishment phases (of endogenous duration) according to a Markov process. If period $t - 1$ was one of successful collusion, then with probability $B = \text{Prob}(p_t < p^+ | I_{t-1} = 1)$ a punishment phase will begin in period t . On the other hand, if a punishment phase began in period $t - 1$ ($I_{t-1} = 0$), the cartel will continue its high output/low price punishment with probability $C = \text{Prob}(p_t \geq p^- | I_{t-1} = 0)$ and will resume successful collusion with probability $D = \text{Prob}(p_t < p^- | I_{t-1} = 0)$.¹⁸ If the industry has been in the punishment state for more than one period, these

¹⁸One of the key additional conditions required for the trigger scheme to have the a simple "tail test" described here is the *monotone likelihood ratio property* (MLRP), defined as

$$\frac{\partial}{\partial p_t} \left(\frac{\partial f / \partial Q_t}{f} \right) < 0,$$

where $f \equiv f(p_t|Q_t)$.

probabilities will be $\text{Prob}(p_t \geq p^- | p_{t-1} \geq p^-)$ and $\text{Prob}(p_t < p^- | p_{t-1} \geq p^-)$ respectively.¹⁹ For a variant of a model with similar qualitative implications, consider the price-secrecy model of Tirole (1988), section 6.7, which predicts price wars as triggered by recessions. In Tirole's model, the firms produce a differentiated product but do not observe their competitors' prices; instead, they try to infer these prices from their own demand. This assumption is in line with Stigler (1964). This model also predicts that the industry switches between collusion and punishment (in this case Bertrand) phases. Hence, in this theory, price wars can be involuntary, in that they may be triggered by an unobservable negative demand shock and not necessarily be attributable to secret undercutting by a cartel member.

7.2 Econometric Testing Framework Adapted to Models of Collusion

For econometric purposes, the following simple model of a symmetric oligopoly nests in a simple way the Rotemberg and Saloner (1986) and the Abreu et al. (1986) game-theoretic models of price fixing outlined in the previous subsection. Suppose that an n -firm price-setting cartel switches between collusive and punishment (or non-collusive) behaviour, according to the collusion indicator function

$$I_t = R_t(\Omega_t), \quad (7.2.1)$$

where Ω_t is the relevant information set available to the firms at time t , and I_t takes the value 0 if punishment (or non-collusive) behaviour occurs in period t , and the value 1 if collusion occurs then.

Further suppose that the industry is characterized by a marginal cost function $MC_t = MC(z_t, \epsilon_{ct})$ and a demand function $q_t = q(p_t, x_t, \epsilon_{dt})$, where z_t and x_t are vectors of exogenous variables, and ϵ_{ct} and ϵ_{dt} are random shocks. Whether or not the realizations of ϵ_c and ϵ_d are observed by the firms at the time of the decision depends on the model under analysis; they are always unobservable by the econometrician. The functions $f(\cdot)$ and $q(\cdot)$ and the exogenous variables are known to the firms. In competitive periods price equals marginal cost, while in collusive periods marginal revenue equals marginal cost. Hence, in the latter periods,

¹⁹Pursuing the implications of this theory further, a positive shift in the distribution of demand due to some exogenous factor would increase the persistence in the Markov process: Imagine a rightward shift of the $f(p_t|Q_t)$ family of distributions, which the firms do not perceive and hence continue employing the same p^+ , p^- thresholds. It is straightforward to see that *both* $A = \text{Prob}(I_t = 1 | I_{t-1} = 1) \equiv \text{Prob}(p_t \leq p^+ | p_{t-1} \leq p^+)$ and $C = \text{Prob}(I_t = 0 | I_{t-1} = 0) \equiv \text{Prob}(p_t \geq p^- | p_{t-1} < p^+)$ will *rise* under such a scenario.

$p_t = MC(z_t, \epsilon_{ct}) - q_t(p_t, x_t, \epsilon_{dt})/q'_t(p_t, x_t, \epsilon_{dt})$, where $q'_t(p_t, x_t, \epsilon_{dt}) \equiv \frac{\partial}{\partial p_t} q_t(p_t, x_t, \epsilon_{dt})$. Let $\mathbf{1}(A)$ be the indicator function, taking the value 1 if logical condition A is true, 0 otherwise. In this notation, $I_t \equiv \mathbf{1}(\text{collusion is effective in period } t)$. The evolution of p_t can thus be summarized by

$$p_t = MC(z_t, \epsilon_{ct}) - I_t \cdot q(p_t, x_t, \epsilon_{dt})/q'(p_t, x_t, \epsilon_{dt}). \quad (7.2.2)$$

For the econometric implementations I assume the following parameterization of the demand function:²⁰

$$q_t = ae^{-p_t/g(x_t, \epsilon_{dt})}. \quad (7.2.3)$$

This functional form is chosen so that price will be independent of x_t in competitive periods but will vary positively with $g(x_t)$ in collusive ones, since $q_t/q'_t = -g(x_t, \epsilon_{dt})$. For tractability, assume further that $g(x_t, \epsilon_{dt}) = \exp(x_t\beta) + \epsilon_{dt}$ and $MC_t = MC(z_t) + \epsilon_{ct}$. I also assume that ϵ_c and ϵ_d are independent of one another, independent over time, and independent of x_t and z_t . The price and (log) quantity equations for observation t in the two regimes can then be shown to be:

Non-Collusive Behaviour:

$$\begin{aligned} I_t = 0 \quad & p_t = MC(z_t) + \epsilon_{p0t} \\ & \ln q_t = \ln a - MC(z_t)/\exp(x_t\beta) + \epsilon_{q0t} \end{aligned} \quad (7.2.4a)$$

Collusive Behaviour:

$$\begin{aligned} I_t = 1 \quad & p_t = MC(z_t) + \exp(x_t\beta) + \epsilon_{p1t} \\ & \ln q_t = \ln a - (1 + MC(z_t))/\exp(x_t\beta) + \epsilon_{q1t} \end{aligned} \quad (7.2.4b)$$

The two sets of price and quantity equation errors, ϵ_{p0} , ϵ_{q0} , ϵ_{p1} , and ϵ_{q1} , are nonlinear functions of the demand and supply errors ϵ_d and ϵ_c , and have the property that $E(\epsilon_{p0}|X, Z) = E(\epsilon_{p1}|X, Z) = E(\epsilon_{q0}|X, Z) = E(\epsilon_{q1}|X, Z) = 0$.²¹

The game-theoretic models under study differ primarily in their implications

²⁰See Roth (1988) for a variant of this model that also allows parametrically different degrees of collusion. The model presented here possesses the property that expected output under collusion is $1/e$ times the competitive one.

²¹Equations (7.2.4a) and (7.2.4b) assume implicitly that

$$E\left(\frac{f(z) + \epsilon_c}{\exp(x\beta) + \epsilon_d} \mid z, x\right) \approx \frac{f(z)}{\exp(x\beta)} + \epsilon_q.$$

Under these conditions, the main requirement for consistency of the ML procedures is satisfied. This specification also implies a specific contemporaneous covariance structure between ϵ_{p0} , ϵ_{p1} , ϵ_{q0} , and ϵ_{q1} in (7.2.4), since, for example, $\epsilon_{p0} \equiv \epsilon_c + \epsilon_d$ and $\epsilon_{p1} \equiv \epsilon_c$. In this paper I neglect the particular form of this covariance structure; ignoring it affects only the efficiency of the estimation procedures.

about the collusion indicator function (7.2.1). Let W_t denote the subset of Ω_t observed by the econometrician, i.e., the set of all relevant exogenous variables in the model. Rotemberg and Saloner (1986) predict that the extent of collusion in some variants of their model, or the probability of a switch into a collusive regime in others, falls as the level of industry demand increases. A simple parameterization of this prediction is

$$I_t = \begin{cases} 1 & \text{if } W_t\gamma + u_t \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (7.2.5.a)$$

Here W_t includes the demand variables x_t in the model, and each element of γ corresponding to one of these variables has the opposite sign that the coefficient of that variable has in the demand function ($\beta_j\gamma_j < 0$). For example, a variable with a positive demand effect ($\beta > 0$) should have a negative effect on the probability of successful collusion ($\gamma < 0$). The error u_t is i.i.d. and (weakly) exogenous with respect to W_t , i.e., $E\{u_t|W_t\} = 0$.

On the other hand, the key prediction of Abreu et al. (1986) is that switching between regimes will evolve according to a Markov process. This theory can be parameterized by

$$I_t = \begin{cases} 1 & \text{if } W_t\gamma + \rho I_{t-1} + u_t \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (7.2.5.b)$$

again with u_t i.i.d. and $E\{u_t|W_t = 0\}$. The Abreu et al. prediction is that the coefficient of ρ should be positive and statistically significant.

7.3 The JEC Data

The data analyzed in this study are 328 weekly observations, from week 1, 1880 to week 16, 1886, of the operation of the Joint Executive Committee (JEC) railroad cartel. As documented by Ulen (1979), the cartel, which primarily shipped grain from Chicago to the East Coast, had extremely varied success in setting price and sharing the market. The effective price charged by each cartel member was not perfectly observable by its rivals because special shipping rates were sometimes secretly arranged with selected customers. The official price of shipping grain, labeled as series *ShipRate*, is plotted in Figure 3. According to MacAvoy (1965), there were two critical periods of non-adherence to the official price: most of 1881 and most of 1884-1885. For detailed historical discussions of the events, see MacAvoy (1965), Ulen (1979), and Roth (1988).

***** FIGURE 3 about HERE *****

The only exogenous information used in the Porter (1983b) and Lee and Porter

(1984) studies was whether or not the lakes were open for navigation. This information is important because lake traffic was the leading substitute for shipping by railway. The dummy variable providing this information is also plotted in Figure 3 as series LakesOpen.

Other important exogenous determinants were left out, however. To remedy this, I compiled and used two additional pieces of exogenous information: an index of extra-cartel railroad competition and an index of total grain produced in the Midwest that might be shipped to the East Coast. For details on the construction of these indices, see Data Appendix. The index of extra-cartel competition, which is plotted in Figure 4a, is based on the simple assumption that the strength of such competition was positively related to the number of railroads that were shipping grain to the East Coast but were operating outside the cartel. The index of total grain production in the Midwest is a value-weighted annual total of the three largest grain crops produced in eight midwestern states, linearly interpolated to obtain weekly values for the index. The total Midwestern grain production index appears in Figure 4b.

*****FIGURES 4a and 4b about HERE *****

Various other sources of information about regime classification are available. Ulen (1983) and MacAvoy (1965) constructed such indicators by relying on perceptions of the effectiveness of the JEC cartel as reported in contemporaneous weekly trade periodicals. A second index of cartel adherence was compiled by Ulen (1979). Porter (1983b) constructed a third index based on the predictions from his econometric model, according to the criterion of maximum estimated probability.²² These regime-classification indicators are incorporated in the econometric implementation of the tests in Section 7.4, using the multiple dummy indicator models developed above.²³

Lack of good technological variables forces me to adopt the further simplifying assumption that, apart from stochastic shocks, marginal cost is constant, or $MC_t = \alpha_0 + \epsilon_{ct}$. This implies we have only demand variables to treat as exogenous, i.e., $W_t = x_t$.²⁴ Four such exogenous variables comprise x_t : a vector of ones, the dummy variable indicating whether or not the lakes were open, the index of extra-cartel railway competition, and the index of grain available to be shipped. *Ceteris paribus*, one expects higher demand for the shipping services of the cartel in periods when the

²²Consider an event with J exhaustive outcomes indexed by $j = 1, \dots, J$, outcome j occurring with probability p_j , $\sum_j p_j = 1$. Denote the predicted probabilities estimated by some model by \hat{p}_j . Then the criterion of maximum estimated probability corresponds to predicting outcome l will occur if $\hat{p}_l = \max_j(\hat{p}_1, \dots, \hat{p}_J)$.

²³These three indices are plotted in Figure 5 of the Online extended version of this paper.

²⁴A simple alternative I plan to explore is that there are seasonal effects in the marginal cost function.

lakes were closed to navigation, when competition from railways outside the cartel was not vigorous, and when there was a big grain crop in the Midwest.

7.4 Empirical Implementation

7.4.1 Estimation Models

To summarize, I use the econometric framework of the switching model of cartel behaviour, presented in Section 2, to test the two game-theoretic models under the additional assumption that marginal cost does not depend on any exogenous information and is constant apart from a random error, i.e., $MC = \alpha_0 + \epsilon_{ct}$:

Non-Collusive Behaviour:

$$I_t = 0 \quad \left\{ \begin{array}{l} p_t = \alpha_0 + \epsilon_{p0t} \\ \ln(q_t) = \ln(a) - \alpha_0 / \exp(x_t \beta) + \epsilon_{q0t} \end{array} \right\} \quad (7.1.0)$$

Collusive Behaviour:

$$I_t = 1 \quad \left\{ \begin{array}{l} p_t = \alpha_0 + \exp(x_t \beta) + \epsilon_{p1t} \\ \ln(q_t) = \ln(a) - (1 + \alpha_0 / \exp(x_t \beta)) + \epsilon_{q1t} \end{array} \right\} \quad (7.1.1)$$

Switching Equation:

$$I_t = \begin{cases} 1 & \text{if } y_{2t}^* = W_t \gamma + \rho I_{t-1} + u_t \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (7.1.2)$$

Coding Error Equation:

$$D_t = \begin{cases} 1 & \text{if } y_{3t}^* = y_{2t}^* + \eta_t \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (7.1.3)$$

In this specification, the parameter α_0 denotes the deterministic part of marginal cost; $\ln a$ and β are demand function parameters (see equation 7.2.3); and γ denotes the effect of the variables W_t on the likelihood of successful collusion in period t . If a variable x_j appears with a positive coefficient β_j , this means that variable x_j has a positive demand impact, *ceteris paribus*. Similarly, a variable w_j which makes collusion more likely should have a positive γ coefficient.

Note that there are two errors for the collusive regime, $\epsilon_1 = (\epsilon_{1p}, \epsilon_{1q})'$, and two for the punishment regime, $\epsilon_0 = (\epsilon_{0p}, \epsilon_{0q})'$. We will be neglecting the correlations between price and quantity errors for a given regime, $\rho(\epsilon_{rp}, \epsilon_{rq})$, $r = 1, 0$ (i.e., $\rho_{01} =$

0). Moreover, the correlations of errors across regimes cannot be identified, hence $\rho(\epsilon_{0i}, \epsilon_{1j})$ will be set to 0, $i, j = p, q$. Finally, also for identification, σ_2 is normalized to 1.

7.4.2 Estimation Results

The estimation results appear in Tables 1–3. The basic model 1NL uses the Ulen (1979) classification of regimes (1 =successful collusion, 0 =price war) and allows for neither a Markov structure nor measurement errors. Model 2NL also uses Ulen’s classification but employs the appropriate methodology of Section 2 to model it as an imperfect scheme. Note that ML estimation of Model 2NL requires the evaluation of bivariate normal integrals.

***** TABLES 1, 2, 3 about HERE *****

The main conclusion that can be drawn from Table 1 is that the decision to treat Ulen’s classification as perfect or imperfect has a serious impact on the estimates. The effect is summarized by a strongly significant variance for the coding error η . Although no coefficient estimate switches sign when imperfections in the regime information are recognized, the predicted regime classifications change substantially. According to the criterion of maximum probability, Model 2NL predicts that 218 out of 328 periods involved collusion, while Model 1NL predicts that in 233 periods the JEC cartel was effective. See Table 4 for the predicted counts from each model and Figure 5 of the Online extended version of this paper for a representation of these predictions.²⁵

The cost parameter is statistically better determined once imperfections in the regime indicators are admitted. My two new exogenous variables are very significant both on the demand side (β coefficients) and in the switching equation (γ ’s). Demand for railroad shipping by the cartel is higher when the lakes are closed, extra-cartel competition is ineffective, and more total grain is available to be shipped. In particular, the importance of lake traffic as a substitute is confirmed as its coefficient is negative and statistically significant. Moreover, contrary to the Rotemberg and Saloner (1986) predictions, the variables that have positive demand effects *raise* the probability that the cartel is colluding effectively. Recall that in the (7.2.5a) parameterization of the Rotemberg and Saloner model, each coefficient corresponding to one of the demand variables entered in the switching equation should have the

²⁵ A third version was also estimated, model 3NL, which combined two sources of regime classifying information using the multiple indicator models developed above. The second regime indicator I tried was the one constructed by Porter (1983b), which employs the predictions from his estimated model. The results from the two-indicator model were very similar to those from the one-indicator model 2NL and are not reported.

opposite sign that the coefficient of that variable has in the demand function. In contrast to this, in all specifications estimated in this paper every demand variable is found to have an effect with the same sign when entered as an exogenous variable in the switching equation. Specifically, the probability of collusion rises with the level of grain available for shipping and falls with effective outside competition and the lakes being open. In his detailed study of the JEC, Ulen (1983) also believes that cartel adherence was strongly positively correlated with cyclical demand conditions; for example, demand upturns appeared to have been enough to extricate the cartel from its 1884-85 price-warfare phase.²⁶

Table 2 shows that the construction of the exogenous variables is probably not inducing serious measurement errors. Allowing for normal errors in the extra-cartel competition index does not affect results substantially, in terms of either the coefficient estimates or the predicted regime classifications.²⁷ An interesting by-product of Table 2 is a comparative evaluation of simulation-based estimation, in particular MSL based on the SRC/GHK simulator. The ML results obtained by numerical quadrature are very close to the MSL estimates when 100 replications were used for MSL.

Table 3 presents the results of estimating the cartel model with Markov structure in the switching equation. The findings in Table 3 lend strong support to the Abreu et al. (1986) model since the coefficient ρ_1 that allows for the first-order Markov structure is very strongly significant. The other parameters are substantially different from those of the (conditional) Bernoulli model, which suggests that coefficient estimates under the (apparently untenable) assumption of Bernoulli switching should not be trusted. Moreover, the models with a Markov structure predict that collusion was effective in fewer periods than suggested by the corresponding models estimated with ρ_1 and ρ_2 set to 0; this finding is more in line with Ulen's regime classification.²⁸ Longer lag structures were tried, but most of the time dependence does not seem to extend beyond two weeks. Some evidence against the optimal one-period punishment

²⁶The cross-country analysis of Suslow (1988) using hazard modelling provides independent confirmation of the finding that successful collusion is more likely to occur in demand booms.

²⁷An issue that remains unanswered is whether incorporating classical measurement errors in the explanatory variables through the approach of Section 4 is too restrictive. As already mentioned, one way to test for this possibility is through nonparametric mixture models. No such attempts are made in this paper, because of the high data requirements of such estimation approaches.

²⁸It is possible that the finding of a strongly significant Markov structure may be caused by residual serial correlation in the unobservables. Unfortunately, even in linear models it is very difficult in practice to differentiate, through the implied common-factor restrictions, the presence of lagged dependent variables as regressors from residual serial correlation. Moreover, given the nonlinearity of the models of this paper, explicit allowance for serial correlation is not feasible with ML methods, because integration of order $T = 328$ would be required.

story can be seen (the asymptotic t -statistic of the coefficient of the second lag being 1.82), but a caveat to be borne in mind is that a week may not be the economically relevant decision-making interval for this cartel.²⁹

It is important to note that other game-theoretic models of oligopolistic behaviour and tacit collusion relax the key assumption made by both Abreu et al. and by Rotemberg and Saloner that demand shocks be *i.i.d.* over time. These models, due to Riordan (1985) and Haltiwanger and Harrington (1988), allow instead for serially correlated demand shocks. It would be interesting to test econometrically whether the two strong findings of this paper, namely existence of a Markov structure in the switching behaviour and the greater likelihood of price warfare in recessions than in booms, would survive such a generalization.

8 Conclusion

This paper discusses switching regressions econometric modelling with imperfect regime classification information. Its econometric novelty is that misclassification probabilities are allowed to vary endogenously over time. Standard maximum likelihood estimation is infeasible in this case because each likelihood contribution requires the evaluation of 2^T terms (where T is the number of observations available). We developed an algorithm that allows efficient estimation when such imperfect information is available, by evaluating the exact likelihood through simply T matrix multiplications (each of a 2×2 matrix times a 2×1 vector.) Our methods were shown to be widely applicable to various areas of analysis such as to Markov-Switching models in Macroeconomics; to external financing problems faced by firms in Corporate Finance; and to game-theoretic models of price collusion in Industrial Organization.

We applied our methods to analyze and test two prototypical game-theoretic models of price fixing and tacit collusion. Switching regressions were used to model the firms' shifting between collusive and punishment behaviour. Our findings cast doubt on the applicability of the Rotemberg and Saloner model to the JEC railroad cartel, while they confirmed the Markovian prediction of the Abreu et al. model of price-fixing behaviour.

²⁹The evidence in Ulen (1979) suggests that all price changes occurred on a Monday. But important time-aggregation issues, of course, remain.

Data Appendix

Constructed Data Series

More details on the construction of these two series can be found in Roth (1988).

1. Construction of Extra-Cartel Railroad Competition Index

The strength of extra-cartel competition as a threat to the JEC cartel is assumed to be directly related to the number of railroads that were outside the cartel and shipping grain to the East Coast. MacAvoy (1965) documents the existence of extra-cartel competing railroads and specifies the exact periods when the JEC cartel responded to the existence of such firms in each case. According to this information, the first extra-cartel competitor was acknowledged by the JEC in week 210 of the sample (January 4, 1884); a second extra-cartel railroad appeared as a competitor on August 15, 1884, which is week 242; finally, a third railroad firm withdrew from the JEC cartel following an unfavourable ruling by a JEC arbitrator and started competing with the cartel in February 6, 1885 (week 267). I follow Roth (1988) and assume that the strength of extra-cartel competition varied with the square-root of the number of railroads operating outside the cartel.

2. Construction of Total Midwest Grain Production Index

Annual data were collected on the largest grain crops (corn, wheat, and oats) from eight Midwestern states. The quantities were weighted according to the average U.S. price for each grain over the period, to generate an annual value index of midwestern grain output. Considerations of a lag between harvest and shipping suggested assigning the annual grain production value to January 1 of the following year. Finally, simple linear interpolation was used to construct weekly values for this index.

References

1. Abreu, D., D. Pearce, and E. Stacchetti (1986): "Optimal Cartel Equilibria with Imperfect Monitoring," *Journal of Economic Theory* 39, 251-269.
2. Berry, S. and H. Briggs (1988): "A Non-Parametric Test of a First Order Markov Process for Regimes in a Non-Cooperatively Collusive Industry," mimeo, University of Wisconsin, Madison.
3. Börsch-Supan, A. and V. Hajivassiliou (1993): "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models," *Journal of Econometrics*, 58(3), 347-368.
4. Cosslett, S. and L.-F. Lee (1985): "Serial Correlation in Latent Discrete Variable Models," *Journal of Econometrics* 27, 79-97.
5. Divgi, D. (1979): "Calculation of Univariate and Bivariate Normal Probability Functions," *Annals of Statistics*, 903-910.
6. Fair, R. and D. Jaffee (1972): "Methods of Estimation for Markets in Disequilibrium," *Econometrica* 40.
7. Fuller, W. (1987): *Measurement Error Models*. New York: Wiley.
8. Goldberger, A. (1972): "Maximum Likelihood Estimation of Regressions Containing Unobservable Variables," *International Economic Review* 13, 1-15.
9. Goldfeld, S. and R. Quandt (1975): "Estimation of a Disequilibrium Model and the Value of Information," *Journal of Econometrics* 3, 325-348.
10. Green, E. and R. Porter (1984): "Non-Cooperative Collusion under Imperfect Price Information," *Econometrica* 52, 87-100.
11. Hajivassiliou, V.A. (1987): "Two Misspecification Tests for the Simple Switching Regression Model," *Economics Letters* 22, 343-348.
12. Hajivassiliou, V. and D. McFadden (1998): "The Method of Simulated Scores for the Estimation of LDV Models," *Econometrica*, July 1998, Vol.66(4), pp.863-896.
13. Hajivassiliou, V. (1993): "Simulation Estimation Methods for LDV Models," in *Handbook of Statistics, Vol.11, Econometrics*, G.S. Maddala, C.R. Rao, and H.D. Vinod, (eds.). Amsterdam: North-Holland.
14. Hajivassiliou, V. (1994): "A Simulation Estimation Analysis of the External Debt Crises of Developing Countries," *Journal of Applied Econometrics*, Vol.9(2), April-June 1994, pp.109-132.

15. Hajivassiliou, V. and F. Savignac (2019): "Novel Approaches to Coherency Conditions in Dynamic Multiperiod LDV Models with State Dependence: With an Application to Financing Constraints and a Firm's Decision and Ability to Innovate," LSE STICERD working paper No.606.
16. Haltiwanger, J. and J. Harrington, Jr. (1988): "The Impact of Cyclical Demand Movements on Collusive Behavior," University of Maryland working paper.
17. Hamilton, J.D. (1996a): "Specification Testing in Markov-Switching Time-Series Models," *Journal of Econometrics*.
18. Hamilton, J.D. (1996b): "This Is What Happened to the Oil Price/Macroeconomy Relation," *Journal of Monetary Economics*.
19. Hamilton, J.D. (1996c): "The Daily Market for Federal Funds," *Journal of Political Economy*, February.
20. Hamilton, J.D. (2001): "A Parametric Approach to Flexible Nonlinear Inference," *Econometrica*, May, vol. 69.
21. Hamilton, J.D. (2002): *Advances in Markov-Switching Models*, Co-edited with Baldev Raj. Physica-Verlag.
22. Hausman, J. (1978): "Specification Tests in Econometrics," *Econometrica* 46.
23. Lee, L.-F. (1978): "Unionism and Wages: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables," *International Economic Review* 19, 415-433.
24. Lee, L.-F. and R. Porter (1984): "Switching Regression Models with Imperfect Sample Separation Information — with an Application on Cartel Stability," *Econometrica* 52, 391-418.
25. MacAvoy, P. (1965): *The Economic Effects of Regulation*. Cambridge: M.I.T. Press.
26. McFadden, D. (1989): "A Method of Estimating Multinomial Probits without Calculating Multiple Integrals," *Econometrica* 57, 995-1026.
27. Moran, (1986): "Orthant Probabilities and Gaussian Markov Processes," *Journal of Applied Probability*, 413-417.
28. Nelson, F. (1977): "Censored Regression Models with Unobservable Stochastic Thresholds," *Journal of Econometrics* 6, 309-328.
29. Pakes, A. and D. Pollard (1989): "The Asymptotics of Simulation Estimators," *Econometrica* 57, 1027-1058.
30. Porter, R. (1983a): "Optimal Cartel Trigger Price Stability," *Journal of Economic Theory* 29, 313-338.

31. Porter, R. (1983b): "A Study of Cartel Stability: The Joint Executive Committee, 1880-1886," *Bell Journal of Economics* 14, 301-314.
32. Riordan, M. (1985): "Imperfect Information and Dynamic Conjectural Variations," *Rand Journal of Economics* 16, 41-50.
33. Rotemberg, J. and G. Saloner (1986): "A Supergame-Theoretic Model of Price Wars during Booms," *American Economic Review* 76, 390-407.
34. Roth, D. (1988): "The Joint Executive Committee Railroad Cartel of 1880-1886: A Reinterpretation and Respecification of Recent Empirical Work," Yale University, mimeo.
35. Steck, G. P. (1958): "A Table for Computing Trivariate Normal Probabilities," *Annals of Mathematical Statistics* 29, 780-800.
36. Stigler, G. (1964): "A Theory of Oligopoly," *Journal of Political Economy* 72, 44-61.
37. Suslow, V. (1988): "Stability in International Cartels: An Empirical Survey," University of Michigan mimeo.
38. Tirole, J. (1988): *The Theory of Industrial Organization*. Cambridge: MIT Press.
39. Ulen, T. (1979): "Cartels and Regulation: Railroad Collusion and the Origins of the Interstate Commerce Commission," unpublished Ph.D. dissertation, Stanford University.
40. Ulen, T. (1983): "Railroad Cartels before 1887: The Effectiveness of Private Enforcement of Collusion," *Research in Economic History* 8, 125-144.

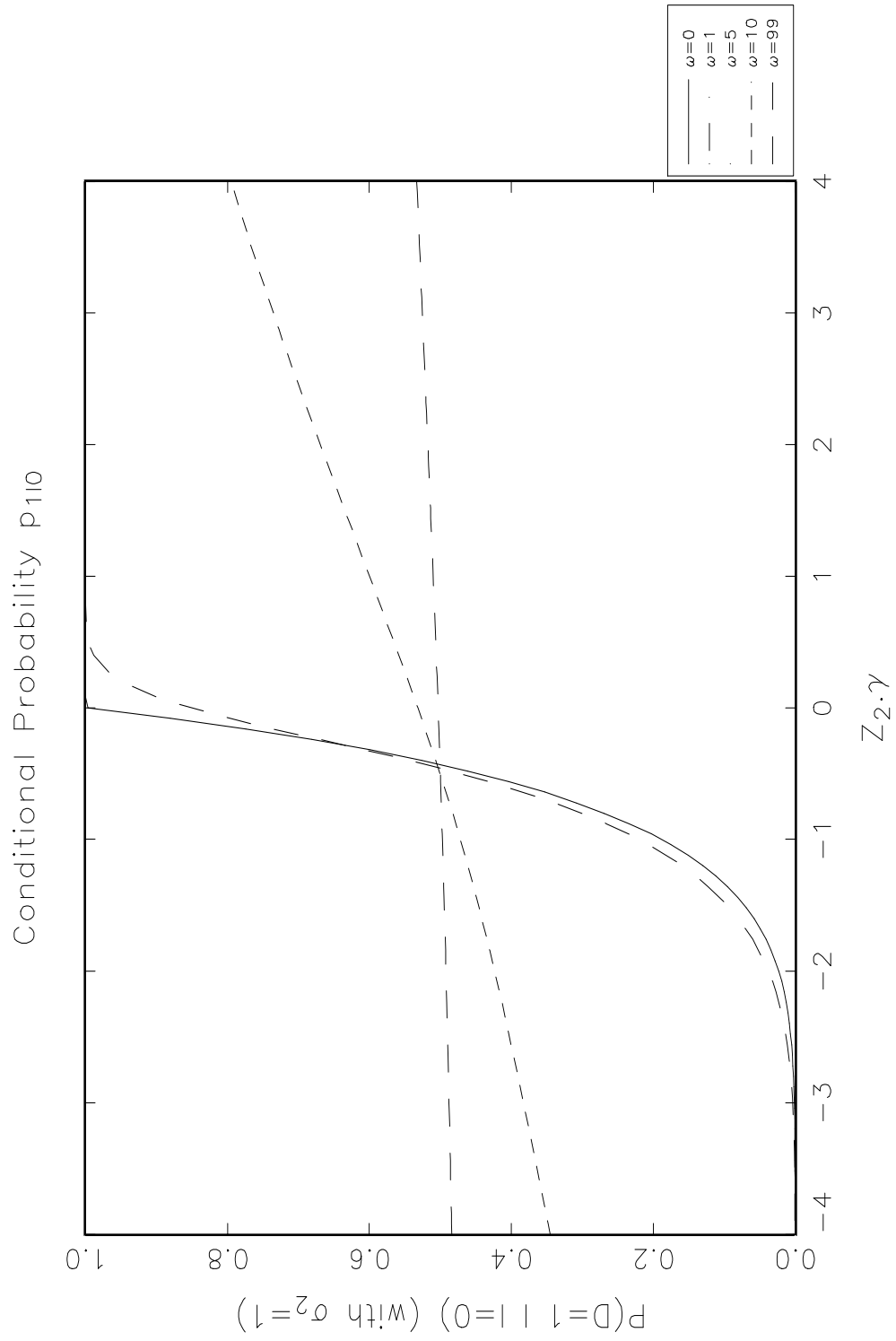


Figure 1a: Conditional Probabilities

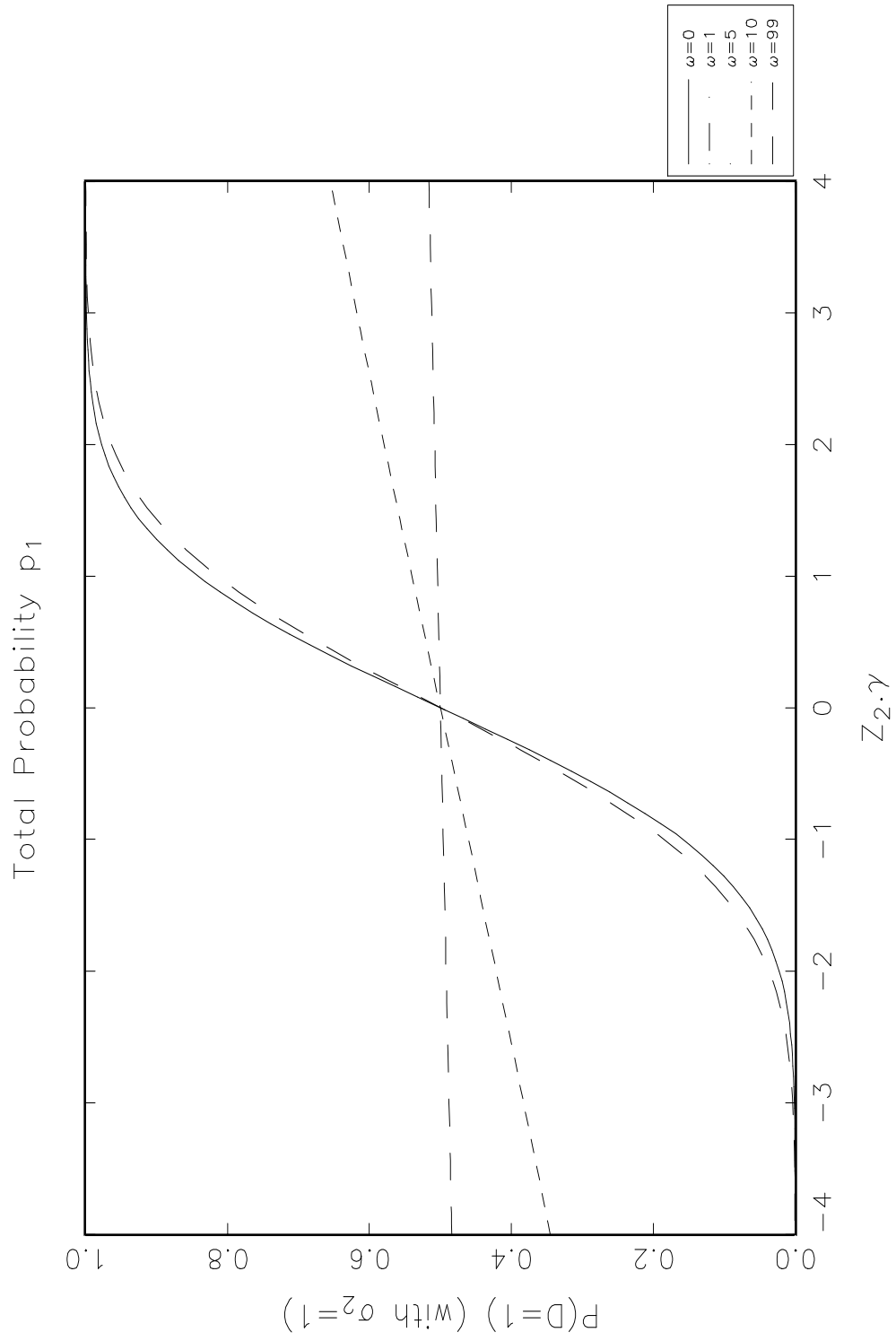


Figure 1b: Total Probabilities

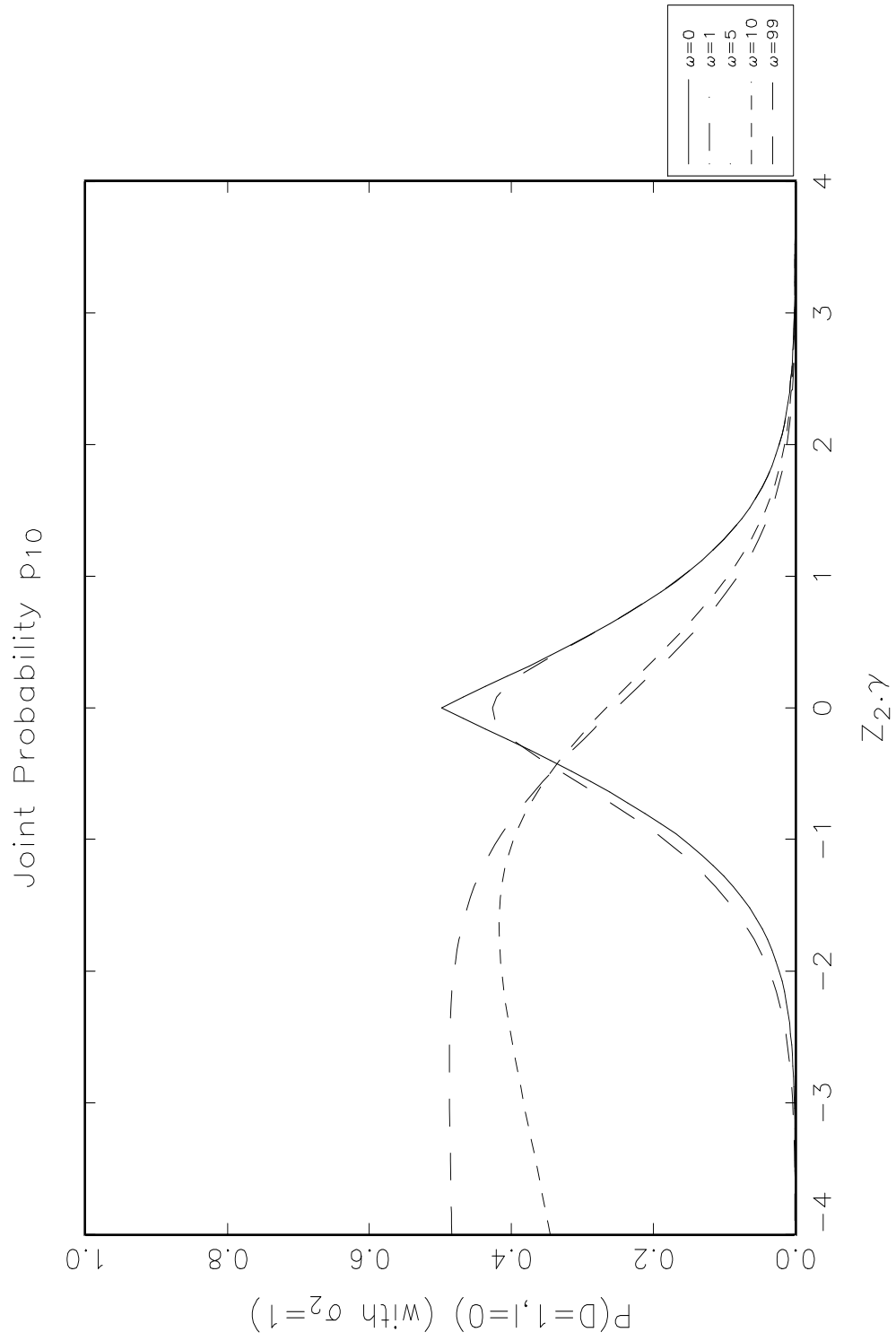


Figure 1c: Joint Probabilities

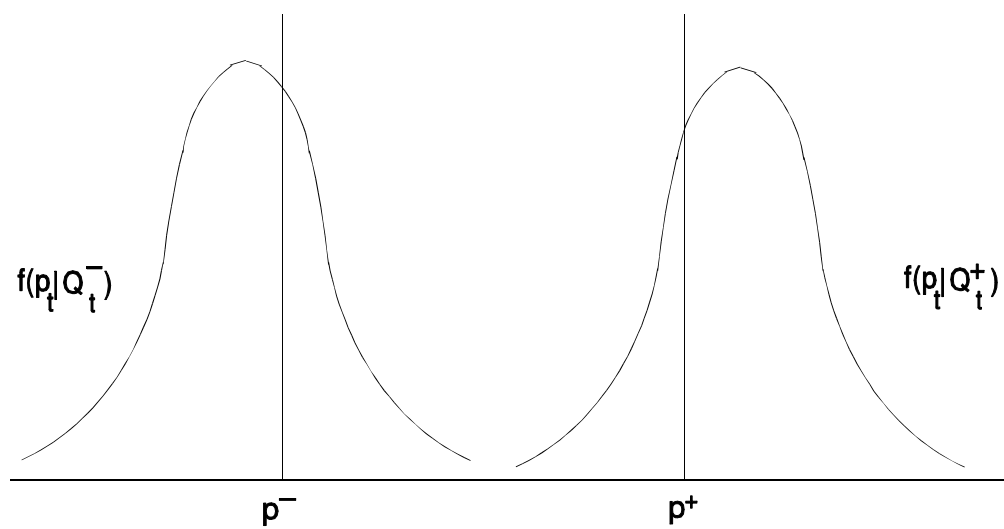


Figure 2: Punishment/Collusion Phases

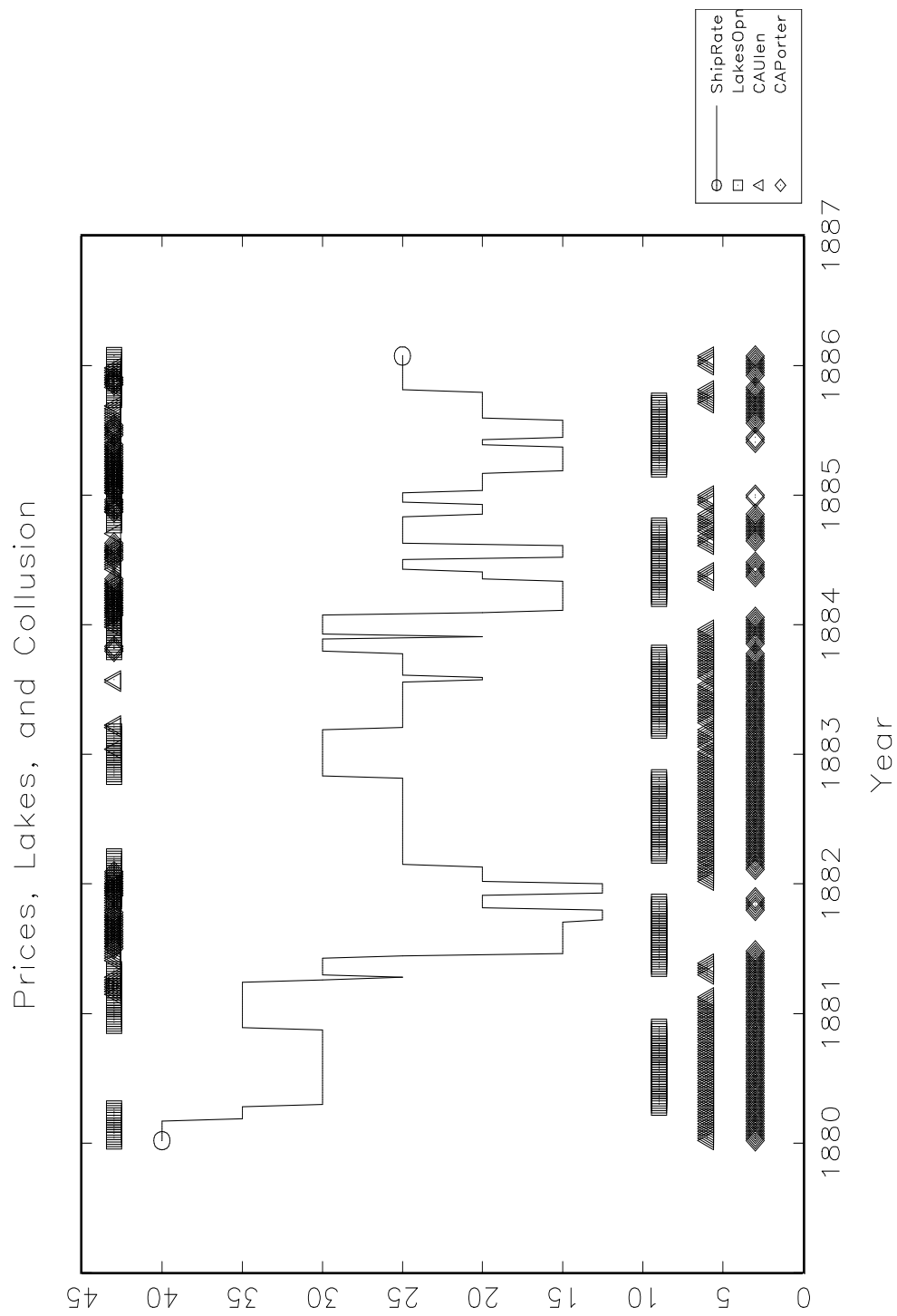


Figure 3: Prices, Lakes, and Collusion

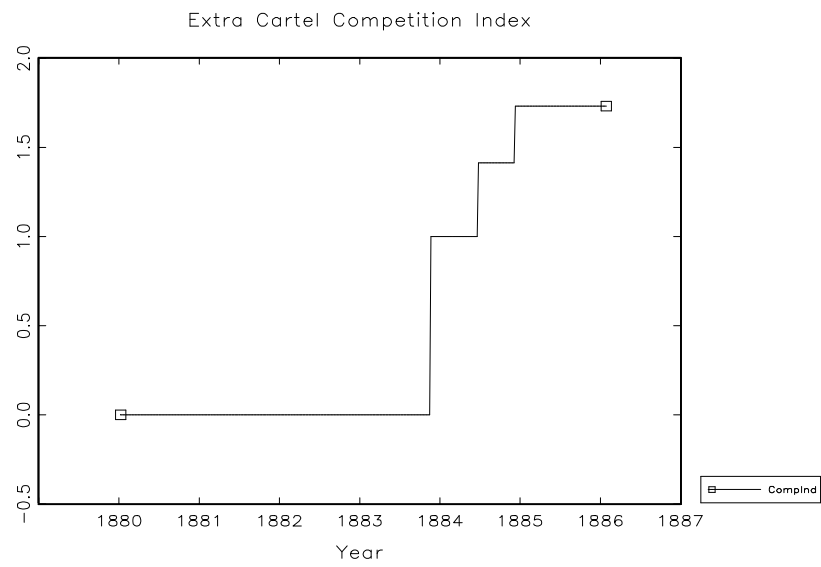


Figure 4a: Extra cartel competition index

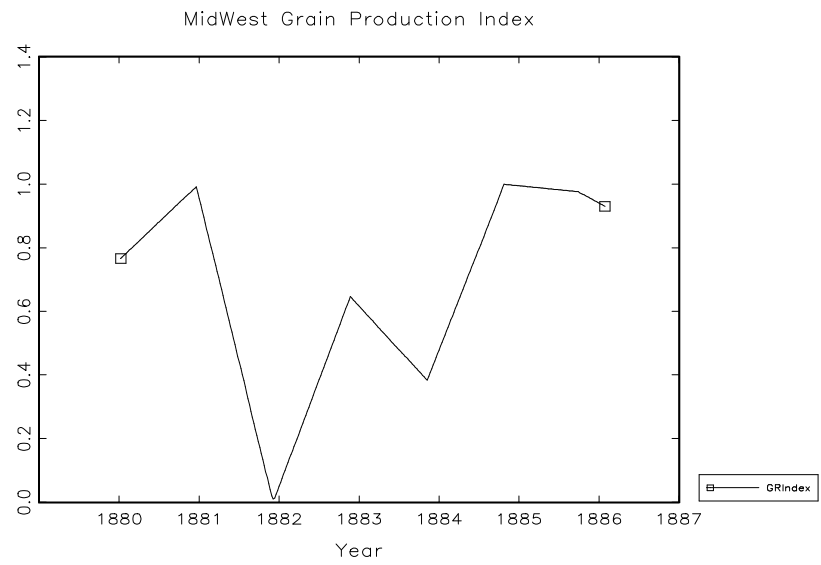


Figure 4b: Midwest Grain Production Index

TABLE 1

BERNOULLI MODELS, NO MEASUREMENT ERRORS IN REGRESSORS

Regime Classification Variable Used: CAUlen (1 = *Collusion*)(Asymptotic *t*-statistics in parentheses)

Variable	Parameter	Model 1NL ¹	Model 2NL ²
	σ_{0p}^2	6.13 (2.59)	5.63 (2.88)
	σ_{1p}^2	5.44 (2.37)	4.27 (2.25)
	σ_{0q}^2	0.27 (2.02)	0.32 (1.76)
	σ_{1q}^2	0.17 (2.01)	0.23 (2.11)
Coding Error	σ_η^2	—	1.63 (3.59)
<u>Marginal Cost Equation:</u>	α_0	16.63 (0.29)	8.21 (2.40)
<u>Demand Equation:</u>	$\ln(a)$	10.96 (0.13)	11.96 (0.68)
	β_0	2.78 (0.11)	3.92 (2.27)
Lakes open dummy	β_1	-0.40 (-10.72)	-0.91 (-12.81)
Midwestern Grain Output	β_2	0.86 (11.12)	0.921 (9.812)
Extra JEC Competition	β_3	-0.44 (-15.53)	-0.34 (-11.43)
<u>Switching Equation:</u>	γ_0	-0.52 (0.32)	-0.72 (-1.95)
Lakes open dummy	γ_1	-0.76 (-2.77)	-0.93 (-4.83)
Midwestern Grain Output	γ_2	5.01 (4.95)	3.78 (7.23)
Extra JEC Competition	γ_3	-2.40 (-5.36)	-3.71 (-7.87)
Loglikelihood		-444.414	-437.342

¹ No lags, no measurement errors² No lags, one imperfect regime indicator

TABLE 2**BERNOULLI MODELS, ERRORS IN EXPLANATORY VARIABLES**Regime Classification Variable Used: CAUlen (1 = *Collusion*)(Asymptotic *t*-statistics in parentheses)

Variable	Parameter	Model 4NL ^{1,2}	Model 5NL ^{1,3}
	σ_{0p}^2	5.23 (3.66)	5.29 (3.06)
	σ_{1p}^2	5.01 (2.58)	4.84 (2.63)
	σ_{0q}^2	0.27 (2.91)	0.44 (3.32)
	σ_{1q}^2	0.21 (2.72)	0.36 (2.89)
<u>Marginal Cost Equation:</u>	α_0	14.63 (0.72)	14.01 (0.63)
<u>Demand Equation:</u>	$\ln(a)$	9.23 (0.27)	9.88 (0.42)
	β_0	2.97 (0.28)	3.21 (0.33)
Lakes open dummy	β_1	-0.77 (-11.25)	-0.88 (-12.38)
Midwestern Grain Output	β_2	1.27 (9.21)	0.71 (8.99)
Extra JEC Competition	β_3	-0.22 (-11.83)	-0.23 (-10.35)
<u>Switching Equation:</u>	γ_0	-0.76 (-0.99)	-0.82 (1.22)
Lakes open dummy	γ_1	-0.53 (-3.67)	-0.59 (-3.17)
Midwestern Grain Output	γ_2	4.83 (3.67)	4.61 (4.81)
Extra JEC Competition	γ_3	-1.83 (-7.48)	-2.01 (-7.12)
Loglikelihood		-456.397	-453.231

¹ Normal measurement errors in Extra Cartel Competition variable² MSM with 100 replications³ ML Quadrature

TABLE 3

MARKOV SWITCHING STRUCTURE,

Regime Classification Variable Used: CAUlen (1 = *Collusion*)(Asymptotic *t*-statistics in parentheses)

Variable	Parameter	Model 6L ¹	Model 7L ²	Model 8L ³
	σ_{0p}^2	6.97 (2.18)	6.78 (2.38)	6.32 (2.42)
	σ_{1p}^2	5.43 (3.36)	5.11 (3.46)	4.37 (2.12)
	σ_{0q}^2	0.97 (3.18)	0.89 (3.27)	0.78 (2.95)
	σ_{1q}^2	0.78 (3.72)	0.81 (3.82)	0.64 (3.47)
<u>Marginal Cost Equation:</u>	α_0	12.63 (0.96)	12.52 (0.95)	10.43 (2.77)
<u>Demand Equation:</u>	$\ln(a)$	7.93 (1.23)	7.83 (1.35)	7.42 (1.08)
	β_0	4.22 (0.74)	4.17 (0.81)	4.12 (0.72)
Lakes open dummy	β_1	-0.23 (-12.78)	-0.28 (-11.81)	-0.35 (-8.42)
Midwestern Grain Output	β_2	0.72 (7.24)	0.77 (7.38)	0.81 (8.15)
Extra JEC Competition	β_3	-0.33 (-12.35)	-0.41 (-11.92)	-0.32 (-9.93)
<u>Switching Equation:</u>	γ_0	-0.50 (0.78)	-0.51 (0.82)	-0.47 (0.58)
Lakes open dummy	γ_1	-0.66 (-3.23)	-0.75 (-3.27)	-0.49 (-3.59)
Midwestern Grain Output	γ_2	3.37 (2.88)	3.42 (2.94)	3.75 (2.73)
Extra JEC Competition	γ_3	-2.93 (-6.72)	-2.74 (-6.84)	-3.02 (-6.32)
Regime Lagged Once	ρ_1	0.78 (6.23)	0.56 (5.72)	0.58 (6.52)
Regime Lagged Twice	ρ_2	—	0.25 (1.82)	—
Loglikelihood		-423.773	-422.661	-418.524

¹ No errors, Markov structure with one lag² No errors, Markov structure with two lags³ One imperfect dummy, Markov structure with one lag

TABLE 4
Regime Classifications:
Indicator Variables and Model Predictions

	Number of Periods in Collusive Regime	Number of Periods in Competitive Regime
Ulen Indicator	203	125
Porter Indicator	238	90
Model 1NL	233	95
Model 2NL	218	110
Model 4NL	231	97
Model 5NL	230	98
Model 6L	208	120
Model 7L	206	122
Model 8L	204	124

Note: All model predictions are according to the criterion of maximum probability defined above.